DOCUMENT RESUME

ED 388 719 TM 024 183

AUTHOR Freedle, Roy; Kostin, Irene
TITLE Semantic and Structural Factors Affecting the Performance of Matched Black and White Examinees on Analogy Items from the Scholastic Aptitude Test.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-91-28
PUB DATE Apr 91
NOTE 43p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Black Students; College Entrance Examinations; *Difficulty Level; Ethnic Groups; High Schools; *High School Students; Item Analysis; Performance; *Racial Differences; Regression (Statistics); Sciences; Semantics; Test Items; Test Results; *White Students
IDENTIFIERS *Analogy Test Items; *Scholastic Aptitude Test

ABSTRACT

Analogy items from the Scholastic Aptitude Test (SAT) were evaluated for differential performance by black and white examinees. Black and white examinees were first matched for overall SAT-V scores prior to conducting item analyses. A content and psycholinguistic analysis of 220 disclosed SAT analogy items from 11 test forms was performed. Regression analyses indicate that black examinees consistently perform differentially better than matched white examinees on the hard analogy items. However, for easy items, particularly those that involve "science" content, white examinees appear consistently to perform differentially better than matched black examinees. In addition, semantic relationships dealing with part/whole relationships in the item stem also contributed negatively to black examinee percent correct responses. Three variables (item difficulty, science content, part/whole relationship) together account for 30% of the variance between the two ethnic (black and white) groups. Of these three significant predictors, two are semantic (part/whole and science content) while the third (item difficulty) reflects a non-semantic factor. Several hypotheses are advanced to explain these findings. Appendix A lists scoring categories, and appendix B lists the variable names and presents a table of intercorrelations of 39 variables. (Contains 2 tables, and 18 references.) (Author/SLD)

RESEARCH REPORT

SEMANTIC AND STRUCTURAL FACTORS AFFECTING THE PERFORMANCE OF MATCHED BLACK AND WHITE EXAMINEES ON ANALOGY ITEMS FROM THE SCHOLASTIC APTITUDE TEST

Roy Freedle
Irene Kostin

Educational Testing Service
Princeton, New Jersey
April 1991

Semantic and Structural Factors Affecting the Performance of Matched Black and White Examinees on Analogy Items from the Scholastic Aptitude Test

Roy Freedle and Irene Kostin

Educational Testing Service

Princeton, NJ

Acknowledgements

This study was supported by Educational Testing Service through Program Research Planning Council funding (PRPC).

We wish to thank Ed Kulick for analyzing DIF values for 9 of our 11 SAT test forms. We also wish to thank N. Dorans and A. Schmitt for providing us with the DIF values for two additional SAT test forms. In addition we also wish to thank L. Schwartz for helping us code some of the semantic relationships in our analogy items.

Abstract

Analogy items from the Scholastic Aptitude Test (SAT) were evaluated for differential performance by Black and White examinees. Black and White examinees were first matched for overall SAT-V scores prior to conducting item analyses. A content and psycholinguistic analysis of 220 disclosed SAT analogy items was performed. Regression analyses indicate that Black examinees consistently perform differentially better than matched White examinees on the hard analogy items! However, for easy items, particularly those that involve "science" content, White examinees appear consistently to perform differentially better than matched Black examinees. In addition, semantic relationships dealing with part/whole relationships in the item stem also contributed negatively to Black examinee percent correct responses. Three variables (item difficulty, science content, part/whole relationship) together account for 30% of the variance between the two ethnic (Black and White) groups. Of these three significant predictors, two are semantic (part/whole and science content) while the third (item difficulty) reflects a non-semantic factor. Several hypotheses are advanced to explain these findings.

Introduction

The purpose of this study is to discover which factors contribute significantly to observed differences between Black and White examinees in their performance on analogy items of the Verbal Scholastic Aptitude Test (SAT). The analogy subtest was focused on here rather than other verbal item types of the SAT because previous research has suggested that, when Black and White examinees are matched on verbal SAT score, Black examinees perform differentially more poorly on the analogy items of the Verbal SAT as compared to the three other kinds of items included in the test (Dorans, 1982).

Numerous possible factors are investigated in this study in order to ascertain whether any could be used to explain differential performance between Black and matched White examinees on analogy items. One factor, the presence or absence of science content in the item, was suggested by a review of the literature (Boldt, 1983). Boldt (1983) found that Black examinees performed differentially more poorly than matched White examinees on verbal items which had science content.

One statistic used to compare Black and White examinees on their performance on SAT analogy items is the DIF value (differential item functioning value) which was developed by Dorans and Kulick (1983). This statistic compares performance on the individual items of the verbal SAT for Black and White examinees who are matched on their total Verbal SAT Scores. In general, a DIF value for an item is computed by assessing the difference between the percent of Black examinees who get the item correct for a given SAT score and the percent of White examinees who get the item correct who also have the same SAT score. When the Black examinees perform differentially more poorly than the matched White examinees on an item, that item obtains a negative DIF value. If the Black examinees perform differentially better than the matched group of White examinees on an item, the item yields a positive DIF value.

A DIF analysis for a given verbal test form yields positive values for approximately half the verbal items while the remaining items yield negative DIF values. This is a consequence of the fact that the sum of all the verbal DIF values will be approximately zero. Each verbal test form includes the analysis of four kinds of verbal items: analogies, antonyms, sentence completions, and reading comprehension. For a given item type there is no a priori reason why, say, all antonyms could not have all negative DIF values, or say, all the sentence completion items might have all positive DIF values. This does not mean that such patterns will actually occur, only that there is no necessary constraint for the DIF values to distribute themselves in any particular pattern within a given item type. All that's required is that all DIF values across a test form sum approximately to zero.

Nevertheless several hypotheses can be advanced regarding the possible distributions of DIF values for verbal analogies.

Some possible patterns for the distribution of DIF values.

1. The expectation originally seemed to be (e.g., Dorans & Kulick, 1983) that only very unusual items would show large positive or large negative DIF values. The additional implication seems to be that the remaining items would be expected to yield DIF values close to zero. The expectation can be further specified by the following consideration. Regardless of which particular semantic or structural characteristic of, say, analogy items is selected for partitioning items, the resultant categories (which stem from the partitioning) should be randomly associated with the occurrence of these extreme DIF values.

The reason this could occur as a likely pattern is that only a few items will have gotten past the scrutiny of several item reviewers (who specifically are looking for items that may favor one group over another). The DIF procedure would be a way of finding those few items that have escaped earlier detection. One would also expect that such highly deviant items might occur anywhere in a set of items. If one selected item difficulty as a relevant way to partition the set of analogy items, one would have no a priori reason to expect that a highly deviant item would be an easy item or a difficult item. That is, there would seem to be little reason to suppose that, say, just easy items might be more likely to yield large positive or negative DIF values than, say, hard items.

2. Another possible pattern that might emerge when comparing minority examinees with White examinees (who typically form the great majority of test takers) is that perhaps the easy items for each of the four verbal item types might show slightly differentially better performance (i.e., positive DIF values) by the minority examinees. If such a pattern occurs, the consequence would be that all hard items would have to yield negative values (since all DIF values must sum approximately to zero). If found, this would in turn imply that the minority test takers are experiencing differentially greater difficulty with just the hard items.

3. Another possibility that could emerge is that easy items for all verbal item types might be differentially more poorly responded to by the minority examinees while all hard items (because all DIF values in a given verbal test form must sum approximately to zero) might be differentially better responded to.

4. Different combinations of some of these above three patterns might also occur. For example, one might find that particular items are occasionally found that are highly deviant (in either a positive or negative direction), but one might still find that they are embedded within a general trend effect such that all easy items, say, have small but negative DIF values while all harder items may have small but positive DIF values, and so on.

Only empirical examination will show which of these several possible patterns is the actual one for any given set of data.

Method

An SAT verbal analogy consists of the following parts:

	Brief description of item parts
Spouse:wife	Item stem
(a) husband:uncle	Incorrect option
(b) son:mother	Incorrect option
(c) child:daughter	Correct option
(d) brother:sister	Incorrect option
(e) grandparent:parent	Incorrect option

Reference to just the item stem will be made below in giving examples of the scoring system.

DIF values were computed, using the standardization method (Dorans, 1982; also see Dorans & Kulick, 1986)[1] for 220 SAT analogy items taken from eleven disclosed SAT forms. Each of these items was coded for the following variables:

1. Item Position - Each SAT form includes two sets of ten analogies each. The number '1' was assigned to the first analogy in each set, '2' to the second, and so on with '10' being assigned to the last member in each set. In each set of ten analogies, the first item is typically the easiest and the tenth item, the hardest.

2. Type of Relation between the Words in the Stem - The relationship between the words in the stem was coded according to a thirteen-category coding system, with some of the categories including a number of subcategories. Altogether (including categories and subcategories), there were twenty-four different codes used in this system. Examples of the types of categories used are part-whole (e.g., forest:tree) and class inclusion (e.g., flower:rose). Using this coding system, two independent coders achieved 72% agreement when coding eighty analogies from four SAT forms. [Percent agreement for part/whole was 96% - this was calculated using a 50-item set containing 36% items coded as part/whole by the more experienced judge; this category is singled out for reasons that will become apparent in the result section.] The reader is referred to Appendix A for the definition of each of the twenty-four relational codes.

In coding this data for analysis each item received a value of '1' for exactly one of these codes and '0' for each of the remainder of these codes.

---

[1]In the version of the formula for computing DIF values used here, examinees who did not reach an item were not included in calculating the percent correct. See Schmitt and Bleistein (1987) for an explanation of why this is the preferred formula to use when comparing Black examinees and White examinees in performance on SAT analogy items.

3. Parts of Speech - Each word in the stem was coded according to whether it was a noun, a verb, or an adjective. Reliability was 100% between two judges for each of these categories as determined by coding 50 words taken from 25 analogy items.

In coding each item for parts of speech we coded each word in the item stem separately. Three columns represented the three possible parts of speech for the first stem word and an additional three columns represented the three possible parts of speech for the second stem word. For example, if the first word of the stem was a noun, we coded a '1' for the noun column representing the first word and a '0' for the adjective and verb columns of the first stem word. If the second word of the stem was a verb it was coded as '1' for another column representing verb use of the second stem word and was also coded as '0' for the remaining two columns for the second word.

4. Abstract versus Concrete - Each word in the stem was coded as either abstract or concrete. A code of '1' was assigned for each concrete word ('0' otherwise). Using a 50 word sample, two judges agreed 96% of the time in coding each word as either abstract or concrete.

5. Animate versus Inanimate - Each word in the stem was coded as either animate or inanimate. A code of '1' was assigned if a particular word was inanimate, '0' otherwise. Using a 50 word sample, two judges agreed 96% of the time in coding each word as either animate or inanimate.

6. Presence or Absence of Science Content - Each analogy stem was coded as to whether or not it contained science content. An example of an analogy stem with science content is: tadpole:amphibian. A code of '1' was assigned if the stem had a "science" content for the word pair; '0' otherwise. Two judges agreed 93% of the time in coding item stems as either science or non-science; 300 items were coded.

7. Presence or Absence of Social/Personality Content - Each analogy stem was coded as to whether or not it contained social/personality type content. An example of an item with such content is: gullible:credulous. If the item had social/personality content it was coded as '1'; '0' otherwise. The judges agreed 92% of the time for 50 item stems in classifying each stem as having social/personality content or not.

8. Frequency of Occurrence - The frequency of occurrence of each word in the stem was obtained from the Francis/Kucera word frequency count (Francis & Kucera, 1982). Actually several derived variables were explored regarding word frequency: the mean frequency of the words in the stem, the log of the more frequent word, the log of the less frequent word, the log of each word frequency, etc. The variable that was the single best predictor was the log of the less frequent word; this is the variable reported below.

All the above variables were correlated with the DIF score using the product moment correlation.

Results[2]

The mean DIF score for the 220 analogy items was: M = -.0051; SD = .0413. A t-test was computed to ascertain whether this mean DIF score differed significantly from a mean of zero--a mean of zero would mean that Black and White examinees, who are matched for verbal SAT scores, did not differ from each other in their overall performance on analogy items.

The result of this t-test [$t(219) = 1.83$, $p < .05$, 1-tailed] indicates that, as shown by the negative mean DIF score, Black examinees overall perform differentially more poorly on analogy items than do White examinees matched on total Verbal SAT scores. [A 1-tailed test is justified here because earlier studies also found that a negative mean DIF value was associated with analogies (Dorans & Kulick, 1926).]

Each verbal SAT form has two sets of 10 analogies. Mean DIF values were computed for the 110 analogy items included in the sets administered first for the 11 SAT forms and also for the 110 analogy items included in the sets administered second. The mean DIF value for the analogy items in the sets administered first was: M = -.0085; SD = .0401. A t-test showed that this mean differed significantly from a mean of zero [$t(109) = 2.34$, $p < .02$]. The mean DIF value for the analogy items in the sets administered second was: M = -.0018; SD = .0423. This latter mean did not differ significantly from a mean of zero [$t(109) = 0.45$, $p > .50$]. Thus, although the mean DIF values for both the first and second sets of analogy items were negative (indicating that Black examinees performed differentially more poorly than White examinees), only the mean DIF value for the sets of analogy items administered first differed significantly from a mean of zero.

Variables Strongly Related to the DIF Value

The variable investigated in this study which has by far the strongest correlation with the DIF value is item position [$r(220) = .502$, $p < .0001$]. Easy analogy items, which occur in lower rank positions, tend to have negative DIF values (Black examinees perform differentially worse than White examinees matched for total Verbal SAT scores); hard analogy items, which occur in higher rank positions, tend to have positive DIF values (Black examinees do differentially better than White examinees matched for total Verbal SAT scores). [In Appendix B we list the predictor variables, present the intercorrelation table of variables and the means and standard deviations of each of the variables.]

Mean DIF values and standard deviations for each item position are presented in Table 1. The t-tests were performed to ascertain whether the mean DIF value for each item position differed significantly from a mean of zero--these results are also presented in Table 1.

---

[2]Many of these results were previously reported in Freedle (1986a) and Freedle (1986b).

Insert Table 1 about here

As we can see from Table 1, analogy items in positions 1, 2, and 4-- typically the easy items--have negative mean DIF values which differ significantly from a mean of zero. [There are 22 items for each of the 10 rank difficulty positions.] Black examinees do significantly worse on these items than do matched White examinees. In contrast, items in positions 7, 8, 9, and 10--typically the hard analogy items--have positive mean DIF values which differ significantly from a mean af zero. Black examinees do significantly better on these harder analogy items than do White examinees with matched Verbal SAT scores!

Although item position clearly had the strongest relationship to the DIF value [$r(220) = .502$, $p < .01$] of all the variables investigated in this study, the following eight variables also yielded significant ($p < .01$) correlations with the DIF value:

Science content, $r(220) = -.328$, $p < .01$;
Social/personality content, $r(220) = .261$, $p < .01$;
First stem word coded as adjective, $r(220) = .230$, $p < .01$;
First stem word coded as noun, $r(220) = -.196$, $p < .01$;
First stem word coded as concrete, $r(220) = -.197$, $p < .01$;
Second stem word coded as concrete, $r(220) = -.236$, $p < .01$;
Stem words have part/whole relationship, $r(220) = -.214$, $p < .01$;
Log of frequency of stem word with lower frequency, $r(220) = -.200$, $p < .01$.

All the eight variables listed above are themselves significantly related to item position. The easier items in the earlier rank positions tend to include nouns, to include concrete words, to have science content and a part/whole relationship between the words, and also to include words with high frequency counts on the Francis-Kucera list. The harder analogy items in the later rank positions tend to include adjectives, to include abstract words, to have social/personality content, and to have a low frequency count on the Francis-Kucera list.

Do any of these variables relate significantly to the DIF value apart from their relationship to item position? Partial correlations were computed to answer this question. Only two variables, i.e., science content and part/whole relationship, remained significantly ($p < .01$) related to the DIF value after their significant relationship to item position had been partialled out..

The above result was also found in the following regression analysis. With DIF value as the dependent variable, and with item position, science content, and part/whole relations as the predictor variables (entered in that order), the following result was obtained.

Table 1

Relationship between Item Position and DIF Scores

Item Position	Mean DIF Value	S.D.	N	t
1	-.0360	.0255	22	6.67**
2	-.0482	.0403	22	5.60**
3	-.0176	.0637	22	1.29a
4	-.0193	.0406	22	2.24*
5	.0006	.0335	22	0.08
6	.0089	.0385	22	1.08
7	.0136	.0291	22	2.19*
8	.0171	.0172	22	4.62**
9	.0107	.0203	22	2.49*
10	.0189	.0170	22	5.25**

* means DIF value is significantly different from 0.0, $p < .05$.
** means DIF value is significantly different from 0.0, $p < .01$.

[a]One item at this position had the most extreme deviancy value, positive or negative, of any of the 220 analogy items. Without this item, the mean DIF score for the remaining 21 items at this position was $M = -.0280$, $SD = .0419$, $t(20) = 3.08$, $p < .01$.

Insert Table 2 about here

In Table 2 we see that the first variable (Item position) accounts for 25.23% of the variance of DIF values. The next variable extracted (science content) accounts for an additional 2.9% of the variance and the third variable extracted (Part/whole relation) accounts for yet an additional 2.44%. Thus each of three predictor variables accounts significantly in predicting DIF value magnitudes.

Supplementary analyses

Some additional analyses were undertaken which help to clarify whether there is any particular problem associated with the fact that the DIF value calculation differentially weights the contribution of each subgroup by how many students fall into any given ability level. Before we explain why this weighting might be a problem in interpreting our main results, let us first present these additional findings which are of interest in and of themselves.

For three of our eleven forms, we divided our Black (and White) examinee samples into two subgroups. The lower scoring Black examinees all obtained a verbal SAT score lower than 350, while a higher scoring Black examinee subgroup obtained scores at or higher than 350. A similar division of the White examinee population was made. [The cutoff at 350 was selected because this divided the Black examinee sample into subgroups of approximately equal numbers.] The percentage of lower scoring White examinees who passed each item was compared to the percentage of lower scoring Black examinees who passed the same items. DIF values were computed for this lower scoring subgroup. Similar DIF values were computed for the higher scoring subgroup as well, using the same procedure. Basically, it was found that for the lower scoring Black examinees, the easier items are still performed differentially less well than their matched White co nterparts, and the harder items are still differentially better responded to. This is exactly the pattern that emerged for the larger sample reported in the sections above. Hence, lower scoring examinees did not show any significantly different pattern than did the total group of Black examinees.

The same basic pattern emerged for the higher scoring examinees as well. Harder items were differentially better responded to by the higher scoring Black examinees (in comparison with higher scoring White examinees) and easier items were more poorly responded to by the same higher scoring Black examinees.

As suggested above, these new results are presented for a very special reason. It is clear that while DIF value calculations do weight the contribution of the minority group as a function of how many individuals fall into a given ability level, such differential weightings have not significantly altered the pattern of our main findings: Black examinees do significantly better on hard analogy items in comparison with their matched White counterparts. Also, the new empirical findings confirm that easy analogy

Table 2

Regression Analyses Using DIF Value
as the Dependent Variable with Three Predictor Variables[a]

Predictor Variable	F(1,216)	Multiple R	R Squared	R-Squared Change
Step 1: Item Position (easy/hard)	50.87	.5023	.2523	.2523
Step 2: Science content in stem	8.06	.5304	.2814	.0290
Step 3: Part/whole relation in stem	7.59	.5529	.3057	.0244

[a] This table shows that after item position is partialled out of the regression (step 1), that science content increases the total variance accounted for in the dependent variable by 2.90 percent; after both item position and science have been partialled out, we see that part/whole semantic relationships in the item stem adds an additional 2.44 percent variance accounted for in the dependent variable. The overall F value which used just these three predictor variables is $F(3,216) = 31.7055$, ($p < .01$). The individual F values reported for each variable above reflects the relative contribution of each of these variables (in the order shown) to this total solution.

items are significantly responded to more poorly by the Black examinees in comparison with their matched White counterparts.

There is another way to examine our data in order to answer the following problem. Do the hard items really contribute significantly to the DIF value analysis or is most of the effect being carried by the easier items? That is, if the minority examinee population which is compared with the matched White examinees show a certain pattern of results for DIF values for the hard items, can one really regard this result as significant since both the minority examinee group and their matched White comparison group tend to get hard items wrong. That is, the hard items may not contribute as much to the SAT scores of these examinee populations as the easy do. But we have just shown that with regard to two different ability levels (above and below 350) of the Black (and the White comparison) examinee groups, we have good reasons for regarding the pattern of under- and over-shooting as a highly replicable finding that is not dependent upon the particular ability levels of the groups being compared.

Another aspect of this problem can be phrased as follows: even though one has isolated three predictor variables [item position (easy/hard), science content, and part/whole relationship] as the best predictors of overall DIF value patterns, might it be the case that these three predictors are operating significantly primarily for just the easy items as opposed to the hard items? If this is true, one implication would be that the three predictor variables should not yield a significant multiple correlation with the DIF values for the hard items, but will yield a significant multiple correlation with DIF values for the easy items. That is, if the hard items are below "threshold" for these minority examinee populations (and for their matched White examinees as well), no set of predictor variables should be found that would significantly correlate with the calculated DIF values obtained for these hard items.

We can quickly answer this by reporting the following multiple correlations [these new analyses are based on all 220 analogies]. We computed the multiple correlation of the DIF values of just the hard items (from rank positions 6, 7, 8, 9, and 10) with the three main predictor variables (item position, science content, and part/whole). We then computed the multiple R of the DIF values of just the easy items (from rank positions 1, 2, 3, 4, and 5) with the same three predictor variables.

The multiple correlation ($N=110$) for predicting DIF values for the hard items was equal to .392. The overall F test for this multiple correlation was $F(3,106) = 6.41$, $p < .01$.

The multiple correlation ($N=110$) for predicting DIF values for the easy items was equal to .401. The overall F test for this multiple correlation was $F(3,106) = 6.78$, $p < .01$.

Clearly, DIF values can be predicted as readily and at the same level of significance for hard items as well as the easy items. Hence, there is no reason to think

that the DIF values obtained for the hard analogy items are any less interpretable than DIF values for easy analogy items.

A final problem to be handled has been raised by some of the work presented by Schmitt & Bleistein (1987) who indicate that there are several ways to calculate DIF values: one formula takes into account the fact that not all examinees get to the end of the verbal sections of the SAT; a slightly different formula retains in the calculation of DIF values those individuals who did not reach a particular verbal item. Clearly, if analogy items were always presented at the beginning or in the middle of a verbal test section, one would not be able to detect any differences in DIF values using these two formulas because speededness effects should only show up for item types presented at the end of verbal sections.

It happens that the SAT introduces some variation into where analogy items are placed within its verbal sections (there are always two verbal sections per test form). For some test forms analogies occur in the middle of the first section and at the end of the second section. Yet other test forms present analogies at the end of the first section and subsequently in the middle of the second section. Clearly if one analyzed the DIF values (which are based on a formula that tries to eliminate those individuals who have not reached an item) from just the set of ten items that were known to occur at the middle of any given test form, such a set should be free of any speededness effects because most if not all examinees will have completed such items; however analyses of DIF values which are known to be solely from the ten item sets that occur at the end of a verbal section might conceivably be sensitive to other variables associated with a speededness effect--that is to say, the population for which DIF values have been calculated for analogies occurring at the end versus the middle of a verbal section is not exactly the same population in both cases. One might well question, therefore, whether there is any systematic effect of the reported DIF values associated with the ends of sections as for the values associated with the middles of sections. In other words, since the population is shifting more unpredictably for the end of section DIF values, there may not be as systematic a relationship between our predictor variables and calculated DIF values for these "speeded" sections.

To examine this possibility we divided our 220 items into two equal halves: one half contained 110 analogy items from only the middle section of the verbal test (the so-called "non-speeded" items) while the remaining 110 items represented DIF values obtained from the end sections (the so-called "speeded" items). We are interested in whether the three most important predictor variables (item difficulty, science content, and part/whole relationship) appear to do as well in predicting the set of DIF values from the end of the test as they do in predicting the set of DIF values from the middle of the test. In particular we found that the multiple correlation for the middle section analogies yielded an F of $F(3,106) = 14.74$, $p < .01$. The percentage of variance accounted for by the three variables was 29.44%. The multiple correlation for the end section analogies yielded an F of $F(3, 106) = 17.21$, $p < .01$. The percentage of variance accounted for by the three variables here was 32.75%. Clearly these two separate analyses show that the predictor variables yielded about the same level of significance

regardless of whether the items were from the so-called "speeded" or "non-speeded" sections of the test.

Discussion

Consistent with previous results, (e.g., Dorans, 1982), we found a significant negative mean DIF value for the 220 analogy items investigated in this study. This finding indicates that, overall, Black examinees perform differentially more poorly than do White examinees, with matched verbal SAT scores, on the set of SAT analogy items. (In this discussion, when we say that a mean DIF value is significant, we mean that it differs significantly from a mean of zero--a mean of zero would indicate no difference in the performance of Black and White examinees matched for Verbal SAT scores.)

Follow-up analyses showed that only the negative mean DIF value for the 110 analogy items in the first sets of analogies administered (Mean = -.0085, SD = .0401) was significant, i.e., $t(109) = 2.34$, $p < .02$. While the mean DIF value for the 110 analogy items in the second sets of analogies administered (Mean = -.0018, SD = .0423) was not significant, i.e., $t(109) = 0.45$, n.s. Thus, Black examinees performed significantly more poorly than White examinees with equal verbal SAT scores on the first sets of analogy items, but on the second sets of analogy items, no overall mean difference in performance was observed between Black examinees and this same group of White examinees.

The above set of findings suggests that Black examinees showed improvement relative to White examinees in their performance on analogy items over the course of the SAT test. One possible explanation for this improvement concerns the different placement of the analogy items in the first versus the second section of the Verbal SAT. For most of the 11 SAT forms used in this study (i.e., for 7 of the 11) the first set of analogies appeared at the end of the first verbal section, while the second set of analogies appeared in the middle of the second verbal section. Perhaps Black examinees do differentially more poorly on the first sets of analogies because the first set of analogies occurred primarily at the end of the section and, for some reason, Black examinees perform more poorly on any kind of item located at the end of a section. To check out this explanation, the mean DIF analogy values were computed for (1) the end of the first verbal section of the seven forms mentioned above, (2) the middle of the second verbal section of these same seven forms, (3) the end of the second section of the remaining four forms (out of 11 forms), and (4) the middle of the first section of these remaining four forms.

About the same amount and direction of change in mean DIF value from the first to the second set of analogies occurred with both kinds of verbal SAT forms (the set of seven and the set of four). Thus, Black examinees' improved performance relative to White examinees on the second set of analogies was not due to the different placement of the analogies in the first versus the second part of the verbal SAT.

Another possible explanation of the pattern of findings for the first versus the second set of analogies is in terms of learning or practice. According to this point of view, Black examinees performed differentially more poorly than White examinees on the first set of analogies because they had less experience with this kind of item as compared to White examinees. This relative lack of experience could be associated with a greater potential for improvement, and thus the practice with the first set of analogies "made up for" the initial lack of experience so that there was no significant difference in performance between Black and White examinees, matched for Verbal SAT, on the second set of analogies.

While we have no direct evidence that Black examinees have less overall experience with analogies than do White examinees, there is evidence (Boldt, Centra, & Courtney, 1986) that Black examinees probably have less experience in taking the SAT (which of course includes exposure to analogies). Also in a recent study conducted with Black and White undergraduates at a local state college [see Freedle, Kostin, & Schwartz (1987)], only 9.4% of the Black students as compared to 58.3% of the White students, took the SAT more than once. This difference is highly significant ($p < .001$). These results suggest that the improvement in Black analogy performance on the second set of analogies within each SAT form might be explained as due to a practice effect.

Variables Which Are Related to the DIF Value

The variable which showed, by far, the strongest relationship to the DIF value was item position: the easy items in earlier rank positions tend to have negative DIF values (indicating differentially poorer performance by Black examinees), while the harder items in later rank positions tend to have positive DIF values (indicating differentially better performance by Black examinees)--also see Freedle and Kostin (1990). This finding is partly consistent with earlier findings which used different methods to assess bias in SAT items. Flaugher and Schraeder (1978), who compared Black and White examinees on item difficulty indices, reported that the easier SAT items were differentially more difficult for Black examinees as compared to White examinees. These authors mention that other studies, which used a scatter plot method to detect bias, also found that Black examinees performed relatively more poorly on the easy items. (In these earlier studies, which did not equate for total SAT score, the harder SAT items were also found to be more difficult for Black examinees as compared to White examinees, but the group difference was not as great as it had been for the easier items.)

The finding reported here using DIF values offers an important new addition to these earlier findings. Unlike the methods used in the earlier studies, this type of analysis first matches Black and White examinees on total SAT score (in the case of this study, total verbal SAT score). When such matching is done, we find that Black examinees, consistent with earlier findings, perform differentially more poorly than White examinees on the easier analogy items; but now, in addition, we find that Black

examinees, matched with White examinees on total verbal SAT score, perform differentially better on the harder items.

It should be noted here that for the SAT analogies, item difficulty is confounded with item position. Thus, for these data we cannot say whether Black examinees perform better on later items because they are harder or because they occur in a later position. (The latter possibility would suggest a "practice effect.") However, in Freedle et al. (1987) item difficulty was not confounded with item position; Blacks still performed differentially better on the hard items and differentially worse on the easy analogy items.

Three possible interpretations will be offered to account for the DIF-value/item-position relationship found in this study. First of all, as was the case with the pattern of findings for DIF values on the first versus the second set of analogies, it may be possible to explain the relationship between item position and DIF value in terms of learning or a practice effect. From this perspective the shift from negative DIF values for the easy analogy items in the earlier positions to positive DIF values for the harder items in the later positions may indicate that Black examinees show more learning or a stronger practice effect over the series of analogy items in comparison to matched White examinees. The assumption here is that Black examinees have had less previous practice or experience with this task; thus, they may have a greater potential for improvement when taking the SAT. Evidence has been presented above supporting the assertion that Black examinees have had less previous experience in taking the SAT as compared to White examinees (Boldt et al., 1986).

A second way of interpreting the DIF-value/item position relationship is in terms of differences between Black and White examinees in cultural background. According to this point of view, Black examinees perform differentially more poorly on the easy items because these items include familiar words which are used frequently in oral conversations at home and with friends and are therefore more susceptible to cultural influences. In contrast, the hard analogy items frequently include difficult words which are learned primarily from books or in academic settings and are rarely used in everyday conversation, e.g., turgidity:nascent. Furthermore, these more difficult words usually have a unique dictionary sense in sharp contrast with the more familiar easy words (see Freedle et al., 1990). The determination of which items will be designated as the "easy" SAT analogy items is based primarily on the responses of the majority White group during item pretesting; these "easy" items, however, will not necessarily be as easy for Black examinees who probably do not share the same cultural background.

There is some evidence that the words most frequently used in oral conversation by Blacks do not completely overlap with the words most frequently used by Whites. Hall, Nagy, and Linn (1984) studied the frequencies with which different words were used in the oral conversation of Black and White preschool children (4.5 to 5.0 years), with each racial group about equally divided by social class--about half middle class and half working class. When the most frequent words used by each racial group were examined, it was found that, although there was a sizeable overlap of words in the "most frequently used" word lists of the Black and White children, there was also a sizeable

vocabulary which was distinctive for each group. Thus, one could conclude that there were many words which were "easy" (i.e., frequently used) for one group but not for the other.

A third interpretation of the DIF-value/item-position relationship is that it is a function of the different content of the easy versus the hard items. The easy items are more likely to have science content, whereas the harder items are more likely to have social/personality content. In order to answer the question of whether the DIF-value/item-position relationship is due to such differences in item content for different item positions, a partial correlation was performed in which the relationship of both science content and of social/personality content to the DIF value was partialled out of the DIF-value/item-position correlation. The resulting correlation, i.e., $r(220) = .42$, $p < .001$, was still highly significant indicating that this DIF-value/item-position relationship was not an artifact caused by differential placement of item content across item positions.

Therefore of the three interpretations, only the last has been ruled out as a viable explanation.

Other Factors Related to the DIF Value

Other variables were found to be significantly related to the DIF value, but none as strongly as item position. Furthermore, all these additional variables were also significantly related to item position. Partial correlations showed that only science content and part/whole relationship remained significantly related to DIF values after their relationship to item position had been partialed out.

Conclusions

The significant negative mean DIF value for the 220 SAT analogy items studied here indicates that Black examinees perform differentially more poorly on analogy items than do White examinees matched for verbal SAT score. Further analyses showed that this difference was primarily due to the Black Examinees differentially poorer performance on the first sets of analogy items (i.e., items included in all the first verbal sections). The DIF value for the second sets of analogies administered did not significantly differ from a mean of zero.

Over and above this shift in mean DIF value across first and second verbal test sections, a further finding in this study is the highly significant relationship between DIF values and item position for SAT analogy items. This correlation indicates that Black examinees perform differentially more poorly on easy analogy items than do White examinees with matched total Verbal SAT score; in contrast, Black examinees perform differentially better than the matched White examinees on hard analogy items. Two interpretations, a "learning" or "practice effect" interpretation and a "cultural difference" interpretation, have been offered to explain this finding. The results of a partial correlation analysis have ruled out an interpretation based on item content.

In general, we raised several possibilities at the beginning of this paper regarding how DIF values might distribute themselves as a function, say, of item difficulty. In particular we suggested that three possible patterns might be (1) a few very large positive and/or very large negative DIF values distributed randomly with respect, say, to item difficulty, (2) generally small DIF values such that easier items tended to yield small positive values while harder items tended to yield small negative values, (3) generally small DIF values such that easier items tended to yield small negative values while harder items tended to yield small positive values.

Of these several patterns, only one is consistent with the findings of the eleven test forms which we have analyzed: the pattern that yields small but generally negative DIF values for the easier analogy items and small but generally positive DIF values for the harder analogy items (also see Freedle & Kostin, 1990). While a few analogy items led to what appear to be large positive DIF values (e.g., dashiki:garment had a positive DIF value of .20 even though it was classified as an easier item--this is about twice the magnitude of the next largest positive DIF value) in general such large values (negative or positive) were of rare occurrence. But the fact that such apparently large DIF values can occur within a distribution of generally smaller positive and negative DIF values shows that a mixed model of the various distributional patterns will probably be needed to develop an adequate statistical model of these types of data.

References

Bejar, I. I., Embretson, S., Peirce, L., & Wild, C. (1984). Applying cognitive research results in GRE test development: analogies. GRE Research Proposal No. 84-19. Princeton, N.J.: Educational Testing Service.

Boldt, R. F. (1983). Status of research on item content and performance an tests used in higher education. Research Report RR-83-3.

Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). The validity of various methods of treating multiple SAT scores. College Board Report #86-4. ETS RR-86-25. Princeton, N.J.

Chaffin, R., & Herrmann, D. J. (1984). The similarity and diversity of semantic relations. Memory and cognition, 12, 134-141.

Dawis, R. V., Soriano, L. V., Siojo, L. R., & Haynes, J. (1974). Demographic factors in the education of relations in analogy word pairs. (Technical Report 3). Minneapolis, Minn.: University of Minnesota, Dept. of Psychology.

Dorans, N. J. (1982). Technical review of SAT item fairness studies: 1975-1979. Unpublished statistical report SR-82-90.

Dorans, N., & Kulick, E. (1983). Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the Standardization approach. (ETS Research Report RR-83-9). Princeton, N.J.: Educational Testing Service.

Dorans, N., & Kulick, E. (1986). Demonstrating the utility of the Standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement, 23, 355-368.

Flaugher, R. L., & Schraeder, W. B. (1978). Eliminating differentially difficult items as an approach to test bias. ETS Research Bulletin RB-78-4.

Francis, W. N., & Kucera, H. (1982). Frequency analysis of English usage: Lexicon and grammar. Houghton Mifflin, Boston.

Freedle, R. (1986a). Second interim report: Report on results of a semantic scoring of four analogy subtests. ETS, unpublished report dated January 10, 1986.

Freedle, R. (1986b). Preliminary findings of two projects dealing with application of a statistic for comparing black and white examinees on SAT analogies, GRE analogies (with additional findings reported for SAT antonym, SAT sentence completion, and SAT reading comprehension items). ETS, unpublished report dated May 16, 1986.

Freedle, R., & Kostin, I. (1990). Item difficulty of four verbal item types and an index of differential item functioning for Black and White examinees. Journal of Educational Measurement, 27, 329-343.

Freedle, R., & Kostin, I., & Schwartz, L. M. (1987). A comparison of strategies used by Black and White students in solving SAT verbal analogies using a thinking aloud method and a matched percentage-correct design. ETS Research Report RR-87-48. Princeton, N.J.

Hall, W. S., Nagy, W., & Linn, R. (1984). Spoken Words: Effects of situation and social group on oral word usage and frequency. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmitt, A., & Bleistein, C. (1987). Factors affecting differential item functioning for Black examinees on Scholastic Aptitude Test analogy items. (ETS Research Report RR-87-23). Princeton, N.J.: Educational Testing Service.

Whitely, S. E. (1977). Relationships in analogy items: a semantic component of a psychometric task. Educational and Psychological Measurement, 37, 725-739.

Wright, D. (1986). An empirical comparison of the Mantel-Haenszel with standardized-methods of detecting differential item performance. Paper presented at AERA in San Francisco, CA, April, 1986.

APPENDIX A

The following list of categories used to score the SAT analogies is a composite of several earlier lists of categories (Dawis, Sioriano, Siojo, & Haynes, 1974; Chaffin & Herrmann, 1984; Whitely, 1977; Bejar, Embretson, Peirce, & Wild, 1984). Also included are some new distinctions which we found necessary to add to the currently avaiiable coding systems. In addition some of the categories used in the earlier systems were dropped by us.

The Semantic System for Scoring Analogies

1. Similarities--synonyms.
The words have similar meanings, such as car:auto, jump:leap, etc. Both words in the analogy are of the same word class (i.e., both are nouns or both are verbs or both are adjectives or adverbs).

2. Similarities--dimensional.
The words are not quite synonyms, but are on the dimension, such as smile:laugh, annoy:torment. (Note: the words differ in magnitude on some putative "intensity" scale.) Both words in the analogy share the same word class.

3. Opposites--antonyms.
The words have opposite meanings, such as happy:sad, alive:dead, etc. Both words share the same word class.

4. Opposites--dimensional.
The words do not quite have opposite meanings, but fall on the opposite ends of a dimension, such as hot:cool, (or laugh:frown). (Note: they differ in two underlying respects: they have some of the antonymic quality and they differ in intensity. Thus "laugh" is a strongly positive quality, while "frown" is a mildly negative quality.) Both words in the analogy are in the same word class.

5. Modifier.
The words are semantically related such that one word is a property or attribute of the otter, such as green:leaf, food:tasty (note that the modifier can be to the right or left of the noun.) One word is an adjective while the other is a noun.

5a. Modifier.definitional (new addition).
Our addition to category 5 is to distinguish between those examples that are necessarily so by virtue of their definition: thus clown-funny expresses not only a modifier relationship but also something that is "definitionally" true. So we would score "sonata:musical" as Modifier.definitional. "Green-leaf" would be scared only as Modifier because a leaf is not by definition green. Other examples of 5a are deleterious:harm, reckless:daredevil.

5b. Negative.modifier (new addition)
Examples are "callous:sensitivity" and "bold:timidity".
Notice that if one changed each noun to an adjective (e.g., callous:sensitive bold:timid) then these would have been scored as antonyms. Since we are using word-class to modulate our semantic distinctions [in the same spirit that Chaffin and Hermann (19: 1) uses case relations to reflect the information conveyed by the syntactic forms] we see that it is critical to try to label each analogy by reflecting not only the semantic senses of the word stems but by taking into account as well their (syntactic) inflections. Thus our semantic categories distinguish some of the following:

bold:timidity	(negative.modifier)
boldness:timid	(negative.modifier)
bold:timid	(antonyms)
boldness:timidity	(antonyms)

Here is an example of a negative.modifier that does not have a true antonymic aspect--anonymous:name. It is clear that anonymous is an adjective while name is a noun. Hence they cannot be antonyms because the word class does not match. But clearly, there is a kind of oppositeness between the words anonymous and name. To capture this aspect we introduced the notion of negative.modifier.

6. Functional.
The words are semantically related such that one word has some function or use for the other. Examples taken from Bejar et al's (1984) list include butcher:cleaver, patron:artist, student:books, car:engine. (Incidentally, we disagree with Bejar's last example "car:engine" since this describes a part-whole relationship--see category "part-whole" described below.) As mentioned above, we decided not to use detailed semantic cases in our scoring system because it is not often possible to identify from just the minimal context of a word pair which case best applies. Because of the paucity of context, it seems ill-advised to make a semantic-case judgment concerning analogies. Instead we choose to score the syntactic form of the analogy (even though this too is sometimes fraught with ambiguity). In practice we think judging syntactic form is more reliable than judging semantic case information.

6a. Functional.definitional (new addition).
"poker:chips" would describe a simple functional relationship; however "poker:cards" is both a functional relationship and additionally is a definitionally true relationship since one cannot play poker without cards. Hence "poker:cards" would be called a "functional.definitional" relationship. Note that the relationship can be asymmetric (and still merit the "definitional" tag) since while every poker game required cards, not all cards are used to play poker.

6b. Negative.functional. (new addition)
If "patron:artist" represents a simple functional relationship, then "patron:hasbeen" would represent a negative functional relationship.

7. Causal.
The words are semantically related such that one is the cause of the other, as in an agent to recipient relationship. An example is bacteria:disease. (Note: with respect to the layperson's knowledge system (not a medical specialist's knowledge system), we normally think of "bacteria" as causing "disease" even though scientifically that is not strictly so.)

7a. Negative.causal (new addition)
Suppose one considers "fungus:decay." This would be a simple causal relationship. But "nonfungus:decay" would represent a substitution of an "antonymic" word for the first member of the pair. Hence we would call "nonfungus:decay" a negative causal whereas "fungus:decay" would just be a simple causal. Another example would be "iodine-:poisoning." This is a simple causal. But "antidote:poisoning" would be a negative causal since an almost antonymic substitution has occurred for the first member of the pair. "Antidote:poisoning" should not be called an antonymic relationship but rather a negative causal; "antidote:poison" however would be called an antonymic relationship--the syntactic word class makes all the difference in what category one places a given analogy. Similarly, "kicking:pain" would be called a simply causal whereas "analgesic:pain" is a negative causal; "happy:painful" would more properly represent a simple antonymic relationship.

7b. Causal.definitional (new addition)
For example, "telescope:magnification", and "mist:dampness". By definition a telescope magnifies and implicitly a "mist" can be damp.

8. Conversion.
The words are semantically related such that in some cases one can become the other, after some process or time lapse or reaction. Examples include grape:wine, colt:horse. [Note: there is some ambiguity associated with the example "colt:horse." If one person visualizes two animals--one a colt and the other a horse--then one wouldn't describe this as a conversion relationship (it would be a "time" relationship according to the Dawis et al. system); but if one visualizes one animal at two different time periods, first as a colt and later as a horse, then conversion is an appropriate description. We mention this issue because it is one of the weaknesses of current scoring systems that contributes to scorer unreliability; a much expanded category system to handle such ambiguities will be undertaken at a later time.] Other examples which we have classified as conversation are corrosion:metal and emendation:text. Here corrosion directly refers to the process by which the second term (metal) undergoes change. Similarly emendation refers to the process by which the second term (text) undergoes change. The first two examples (e.g., grape:wine and colt:horse), while also categorized as conversion, does not directly refer to the process. Fermentation is the precess by which grapes are converted into wine, and development is the process by which a colt is converted into an adult horse.

8a. Negative.conversion (new addition)
E.g., indecipherable:translation
A translation can be considered a deciphering of one language into another. Hence it is a

conversion process. A negative conversion would then be indicated by indecipherable:translation.

9. Action.
The words are semantically related such that one is an action associated with the other. Either agent-action, action-recipient, or action-instrument relationships are included. Examples are knife:cuts, predator:hunts, drink:cup. (Note: we would modify this relationship using category 9a below.)

9a. Action.definitional (new addition)
"Scissors:cutting" does describe an action performed by an instrument (scissors). However, by definition scissors are used for cutting. This definitional quality is missing from a pure action example such as knife:stab. A knife by definition is used to cut, but stabbing is a non-definitional use of knife. Hence by our expanded category system "knife:stab" represents action, "knife:cut" and "scissors:cutting" would be better described as action.definitional.

9b. Action.causal (new addition)
E.g., subjugate:obedience, burnish:luster, net:snare. One of the words is a verb (subjugate) so "action" is a relevant category. Also the action in some sense causes or leads to the second word: so subjugation can lead to obedience, or the action of burnishing can lead to a luster.

9c. Action.causal.definitional (new addition)
E.g., refrigerate:cool, stomach:digest, oil:lubricate. One word is a verb so action is indicated. The other word is causally connected to the action; in addition it is definitionally connected. The action of "refrigerate" necessary leads to cooling by definition.

9d. Negative.action (new addition)
E.g., bluff:intention. The verb "bluff" indicates action is relevant, but a true intention is not exhibited by bluffing, instead bluffing exhibits a fake intention.

9e. Negative.action.causal (new addition)
E.g., fetter:mobility, babble:meaning.
The verb "fetter" indicates an Action code is relevant. Something unfettered makes it more mobile. But "fetter" is the opposite of "unfettered" hence a negative.action.causal code is indicated.

9f. Action.conversion (new addition)
E.g., decipherable:decoded. When one of the words is a verb (decoded) and it also refers to a process which has the effect of "converting" or changing the other word, e.g., magnify:size; assuage:anguish; refine:petroleum; ossify:bone; defame:reputation.

9g. Negative.action.conversion (new addition)
E.g., indecipherable:decoded The single verb "decoded" allows us to select "action" as

relevant; next we note that something that is "decipherable" has been "decoded" and so a conversion or process has taken place. We also note that "indecipherable" is the opposite or negative of "decipherable" so the final scoring is "negative.action. conversion".

10. Class inclusion.
One word names a class that includes the other, such as flower:rose, crime:theft. (Note: it would be totally redundant for us to try to distinguish an additional subclassification such as "class inclusion.definitional" since in every instance of "class inclusion" a definition is implied; for example, in "flower-rose," "rose" is necessarily by definition a flower; and in "crime-theft," "theft" is necessarily by definition a crime. Hence all we need to fully describe this category is the term "class inclusion."

11. Part/whole.
One word names a part of the other, such as link:chain, forest:tree, cow:herd.

lla. Part/whole.definitional (new addition).
We would like to modify Bejar et al.'s (1984) list of examples for part/whole by assigning them to another subcategory called part/whole.definitional. "Forest:tree" exhibits a part/whole relationship, but it also exhibits a definitional quality since a forest necessarily consists of trees; by contrast "building:annex," while it exhibits a part/whole relationship does not exhibit any definitional characteristics (building does not have to have an annex in order to be called a building). "Link:chain" should also be classified as "part/whole.definitional" since a chain necessarily consists of links.

llb. Negative.part/whole. (new addition)
If "cow:herd" is a positive instance of part/whole then "maverick:herd" is an instance of negative part/whole since an animal that is a "maverick" used to be part of the herd but is no longer. Similarly, if "member:society" describes a part/whole relationship then "pariah:society" describes a negative part/whole relationship since a "pariah" is an outcast of society. The category "antonym" by itself is not subtle enough to capture this distinction.

12. Class membership.
The two words are members of the same class, such as dog:bird (pets), fork:tablespoon (utensils).

12a. Class membership.associational. (new addition)
We disagree with the examples given in the Bejar et al.'s (1984) list for this category such as dog:cat (pets), and fork:knife (utensils). We prefer to assign dog:cat and fork:knife to a slightly different subcategory called "class membership.associational." Dog:cat are commonly associated as words in a word-association test, whereas dog:bird would be a rare association. Similarly, fork:knife are common associates whereas fork:tablespoon would not be.)

13. Quantitative.
The words have a relationship of magnitude or number such as dime:dollar, inch:foot. We have restricted this category to word pairs where each word has a clearly defined quantitative attribute. Clearly as these examples illustrate, other distinctions can simultaneously operate within the quantitative distinction, such as part/whole. definitional, etc.

Special note: to be systematic we might have considered generating a negative subcategory for each of the major categories. Also to be complete we might have considered inserting a definitional subcategory for each major category along with a negative/definitional subcategory. However, when no examples were encountered, we did not formulate these additional subcategories. The following subcategories, therefore, will not be included in the augmented scoring system until clear examples are found (some of the below are not in the category system simply because they are redundant with other categories already present):

negative modifier.definitional
negative cause.definitional
negative conversion.definitional
negative action.definitional negative class inclusion
class inclusion.definitional (this is redundant)
negative part/whole.definitional
class membership.definitional (this is redundant)
negative class membership
negative class membership.definitional
negative quantitative
quantitative.definitional (this is redundant)
negative quantitative.definitional
negative (pattern) non-semantic
non-semantic.definitional
negative non-semantic.definitional

General comments on scoring:

To score these categories, the scorer may not change the word class of one word in the pair as in altering a noun to a verb. However, the only exception to this is that both words may be given a different common form in order to facilitate categorization. For example, gobble:eat. You can say that "gobbling" is a form of "eating" when deciding to code this as class inclusion. Both verbs in the word pair have been converted to gerunds (gobble becomes gobbling, eat becomes eating). Or one could have made both verbs into infinitives ("to gobble" is more specific than "to eat").

Sometimes a missing noun has to be filled in (missing case relation) in order to code an analogy. e.g., reprehensible:blame one can expand this to "a reprehensible act can lead to blame". hence "causal". One could determine that reprehensible is causally

connected to blame only by inserting the missing noun (e.g, "act") in order to make the causality more obvious.

Reliable coding is greatly facilitated by noting the following restrictions between semantic category and the syntactic word class of each word in an analogy pair:

a) For categories 1, 2, 3, or 4:
In order to code an analogy as belonging to categories 1, 2, 3 or 4 both words in the analogy pair must have the same word class; that is, both must be nouns, or both must be verbs, or both adjectives (other word classes, such as adverbs, seem to never be used as pairs).

b) For category 5 (modifier):
In order to code an analogy as category 5 (modifier), one of the words has to be an adjective and other a noun; that is, it must be either noun:adjective or adjective:noun. (the directional difference is ignored in this particular scoring scheme).

c) For category 6 (functional):
In order to consider using category 6 (functional) one of the following word class conditions must be met:

1) noun:noun
2) adjective (missing noun):noun
For example, "arable(land):farmers" or
"habitable(land):occupants."

d) For category 9 (action):

In order to consider using category 9, one of the following word class combinations must be present:

1) adjective:verb or verb:adjective
2) noun:verb or verb:noun
3) adjective:adverb or adverb:adjective
4) noun:adverb or adverb:noun

Although certain categories require certain word class combinations, the reverse is not true--certain word class combinations can be associated with more than one category; e.g., noun:noun could be coded in several categories including 1, 6, 10, etc.

If it is not clear whether a word is a noun or a verb (when considering just a single word pair), it is useful to scan the remaining alternatives in an analogy item to help decide the word class. For example, by itself the word "pirouette" in the word pair "pirouette:dancer" could be a verb or a noun. But by scanning the response alternatives such as touchdown:referee, motivation:coach, somersault:acrobat, model:sculptor, rink:skater, it is clear that "pirouette" is to be considered a noun.

We have not used the case frames of Chaffin and Hermann (1984) in any explicit way mainly because they were already implicitly contained in our other categories. We found it more economical to construct a few new categories, such as noting the necessity of adding a missing noun, without additionally stating whether the missing noun was an instrument, an actor or a benefactive. Because there were so few examples (approximately 10 out of 600) that required this "missing noun" category (and because subsequent analyses showed that this missing noun category did not relate to either item difficulty or deviancy scores), we decided that constructing subcategories to reflect benefactive or actor or instrument was unnecessary at this point in our study. Future work of course may require these additional subcategory distinctions to be added to our existing semantic system.

Appendix B

List of Variable Names with their Means and Standard Deviations and a Table of Intercorrelations of 39 Variables

I. Means and standard deviations of 39 variables.

Variable Label	Name of Variable	Mean	S.D.	No. Cases
v1	DIF value	-.0051	.0413	220
v3	Adjective:1st stem word	.1909	.3939	220
v4	Noun:1st stem word	.6136	.4880	220
v5	Verb:1st stem word	.1955	.3975	220
v6	Adjective:2nd stem word	.0500	.2184	220
v7	Noun:2nd stem word	.8682	.3391	220
v8	Verb:2nd stem word	.0818	.2747	220
v9	Category-1 (similarities:synonyms)	.0045	.0674	220
v10	Category-2 (similarities:dimensional)	.0545	.2276	220
v11	Category-3 (opposites:antonyms)	.0091	.0951	220
v12	Category-4 (opposites:dimensional)	.0000	.0000	220
v16	Category-8 (conversion)	.0227	.1494	220
v18	Category-10 (class-inclusion)	.0955	.2945	220
v20	Category-12 (class-membership)	.0045	.0674	220
v21	Category-13 (quantitative)	.0045	.0674	220
v23	Category-5b (negative.modifier)	.0500	.2184	220
v25	Category-6b (negative.functional)	.0273	.1632	220
v26	Category-7a (negative.causal)	.0227	.1494	220
v27	Category-8a (negative.conversion)	.0000	.0000	220
v29	Category-9d (negative.action)	.0409	.1985	220
v31	Category-11b (negative.part/whole)	.0045	.0674	220
v32	Category-12a (class membership.associational)	.0000	.0000	220
v33	Item Position (item difficulty)	5.5000	2.7880	220
v34	Science content	.2182	.4140	220
v39	Category-9e (negative.action.causal)	.0182	.1339	220
v40	Category-9f (action.conversion)	.0000	.0000	220

v41 Category-9g (negative.action.conversion)	.0045	.0674	220
v42 Animate/Inanimate:1st stem word	.8318	.3749	220
v43 Animate/Inanimate:2nd stem word	.7455	.4366	220
v44 Abstract/Concrete:1st stem word	.4273	.4958	220
v45 Abstract/Concrete:2nd stem word	.5273	.5004	220
v137 Log of frequency of less frequent word in stem	.5704	.5191	220
v141 Social/Personality	.3364	.4735	220
v158 Categories 5 plus 5a (modifier)+(modifier.definitional)	.0682	.2526	220
v159 Categories 6 plus 6a (functional)+ (functional.definitional)	.2545	.4366	220
v160 Categories 7 plus 7b (causal)+(causal.definitional)	.1045	.3067	220
v161 Categories 9 plus 9a (action)+(action.definitional)	.0773	.2676	220
v162 Categories 11 plus 11a (part/whole)+ (part/whole.definitional)	.0636	.2447	220
v163 Categories 9b plus 9c (action.causal)+ (action.causal.definitional)	.0682	.2526	220

Note: all of the categories that originally distinguished between a definitional subcategory and a main category (e.g., part/whole and part/whole.definitional which were originally separate categories) were later combined into a single category because we found that while the two categories are cognitively distinct, it was nevertheless difficult to get agreement between two judges as to which of the two categories a particular item belonged--by combining all definitional subcategory and the main category, this difficulty in reliability was avoided. This combination is apparent for variables v1[illegible] through v163 above.

II. Table of intercorrelations of 39 variables.

Below are presented the intercorrelations of 39 variables (vl is the criterion or dependent variable, the remaining 38 are predictor variables). 99.99 in the table means that a correlation could not be calculated due to insufficient variance of the variable. Each variable in this table has been named at the beginning of this appendix (appendix B).

V1	1.00000	0.23019	-0.19638	0.01299	0.10659	-0.11546	0.05774	0.00678	-0.01090	0.00263	99.00000	-0.07580
V3	0.23019	1.00000	-0.61217	-0.23942	0.10083	-0.15260	0.10818	-0.03282	0.03611	0.07533	99.00000	0.00353
V4	-0.19638	-0.61217	1.00000	-0.62116	-0.11779	0.24271	-0.20590	0.05362	-0.09716	-0.02235	99.00000	0.05837
V5	0.01299	-0.23942	-0.62116	1.00000	0.04470	-0.14678	0.14561	-0.03331	0.08351	-0.04721	99.00000	-0.07516
V6	0.10659	0.10083	-0.11779	0.04470	1.00000	-0.58876	-0.06848	-0.01550	0.22042	0.19777	99.00000	-0.03499
V7	-0.11546	-0.15260	0.24271	-0.14678	-0.58876	1.00000	-0.76609	0.02633	-0.32058	-0.10425	99.00000	0.05942
V8	0.05774	0.10818	-0.20590	0.14561	-0.06848	-0.76609	1.00000	-0.02017	0.22041	-0.02859	99.00000	-0.04552
V9	0.00678	-0.03282	0.05362	-0.03331	-0.01550	0.02633	-0.02017	1.00000	-0.01623	-0.00647	99.00000	-0.01030
V10	-0.01090	0.03611	-0.09716	0.08351	0.22042	-0.32058	0.22041	-0.01623	1.00000	-0.02301	99.00000	-0.03663
V11	0.00263	0.07533	-0.02235	-0.04721	0.19777	-0.10425	-0.02859	-0.00647	-0.02301	1.00000	99.00000	-0.01461
V12	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	1.00000	99.00000
V16	-0.07580	0.00353	0.05837	-0.07516	-0.03499	0.05942	-0.04552	-0.01030	-0.03663	-0.01461	99.00000	1.00000
V18	0.03936	-0.15780	0.16246	-0.04309	-0.07453	-0.01060	0.07234	-0.02195	-0.07803	-0.03112	99.00000	-0.04954
V20	-0.07855	-0.03282	0.05362	-0.03331	-0.01550	0.02633	-0.02017	-0.00457	-0.01623	-0.00647	99.00000	-0.01030
V21	-0.12122	-0.03282	0.05362	-0.03331	-0.01550	0.02633	-0.02017	-0.00457	-0.01623	-0.00647	99.00000	-0.01030
V23	0.08785	0.41922	-0.24629	-0.11308	0.04306	0.02774	-0.06848	-0.01550	-0.05510	-0.02197	99.00000	-0.03499
V25	0.03510	-0.01033	0.07555	-0.08253	-0.03841	0.06525	-0.04998	-0.01131	-0.04022	-0.01604	99.00000	-0.02553
V26	0.06049	0.08113	-0.00427	-0.07516	-0.03499	0.05942	-0.04552	-0.01030	-0.03663	-0.01461	99.00000	-0.02326
V27	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V29	0.06698	-0.04193	-0.26028	0.36115	-0.04738	0.01264	0.02207	-0.01396	-0.04961	-0.01978	99.00000	-0.03150
V31	0.00678	-0.03282	0.05362	-0.03331	-0.01550	0.02633	-0.02017	-0.00457	-0.01623	-0.00647	99.00000	-0.01030
V32	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V33	0.50229	0.42682	-0.35914	0.01796	0.07624	-0.16139	0.13857	-0.05882	0.00000	0.00000	99.00000	0.02655
V39	0.06323	0.10702	-0.10163	0.01872	-0.03122	-0.24867	0.33175	-0.00920	-0.03269	-0.01303	99.00000	-0.02075
V40	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V41	0.06253	0.13911	-0.08516	-0.03331	-0.01550	-0.17342	0.22637	-0.00457	-0.01623	-0.00647	99.00000	-0.01030
V34	-0.32813	-0.17260	0.23836	-0.12161	-0.12119	0.14078	-0.07739	-0.03570	-0.12689	-0.05060	99.00000	0.21483
V42	0.05228	0.21842	-0.35679	0.22163	0.04740	-0.10336	0.08989	-0.15028	0.10800	-0.08497	99.00000	-0.17606
V43	0.04248	0.17765	-0.18508	0.05119	0.13406	-0.22770	0.17443	-0.11564	0.14036	-0.05397	99.00000	-0.12094
V44	-0.19696	-0.41956	0.68536	-0.42572	-0.15599	0.22791	-0.15726	0.07923	-0.08607	0.01408	99.00000	0.05325
V45	-0.23630	-0.30452	0.35187	-0.13024	-0.24229	0.41152	-0.31526	0.06398	-0.13340	-0.00523	99.00000	0.08331
V141	0.26069	0.14376	-0.10688	-0.01125	0.05739	-0.06386	0.03319	-0.04811	0.12555	0.03317	99.00000	-0.10857
V137	-0.25993	-0.25711	0.19703	0.01289	-0.08061	0.09895	-0.05803	0.17023	0.00822	-0.10548	99.00000	-0.05121
V158	0.09415	0.41922	-0.22979	-0.13333	0.18617	-0.05452	-0.08075	-0.01828	-0.06497	-0.02591	99.00000	-0.04125
V159	-0.06604	-0.20420	0.39938	-0.28802	-0.13406	0.22770	-0.17443	-0.03949	-0.14036	-0.05597	99.00000	-0.08911
V160	0.00470	-0.05258	0.14909	-0.13095	-0.07839	0.13314	-0.10200	-0.02309	-0.08207	-0.03273	99.00000	-0.05211
V161	-0.01892	-0.14057	-0.29478	0.50127	-0.06639	0.01212	0.03783	-0.01955	-0.06951	-0.02772	99.00000	-0.04413
V162	-0.21397	-0.12663	0.20686	-0.12849	-0.05981	0.10158	-0.07782	-0.01762	-0.06262	-0.02497	99.00000	-0.03976
V163	0.06174	-0.03963	-0.26683	0.36691	0.18617	-0.26775	0.18243	-0.01828	-0.06497	-0.02591	99.00000	-0.04125

	V19	V20	V21	V23	V25	V26	V27	V29	V31	V32	V33	V39
V1	0.03936	-0.07855	-0.12122	0.08785	0.03510	0.06049	99.00000	0.06698	0.00678	99.00000	0.50229	0.06323
V3	-0.15780	-0.03282	-0.03282	0.41927	-0.01033	0.08113	99.00000	-0.04193	-0.03282	99.00000	0.42682	0.10702
V4	0.16246	0.05362	0.05362	-0.24629	0.07555	-0.00427	99.00000	-0.26028	0.05362	99.00000	-0.35914	-0.10163
V5	-0.04309	-0.03331	-0.03331	-0.11308	-0.08253	-0.07516	99.00000	0.36115	-0.03331	99.00000	0.01796	0.01872
V6	-0.07453	-0.01550	-0.01550	0.04306	-0.03841	-0.03499	99.00000	-0.04738	-0.01550	99.00000	0.07624	-0.03122
V7	-0.01060	0.02633	0.02633	0.02774	0.06525	0.05942	99.00000	0.01264	0.02633	99.00000	-0.16139	-0.24867
V8	0.07234	-0.02017	-0.02017	-0.06848	-0.04998	-0.04552	99.00000	0.02207	-0.02017	99.00000	0.13857	0.33175
V9	-0.02195	-0.00457	-0.00457	-0.01550	-0.01131	-0.01030	99.00000	-0.01396	-0.00457	99.00000	-0.05882	-0.00920
V10	-0.07803	-0.01623	-0.01623	-0.05510	-0.04027	-0.03663	99.00000	-0.04961	-0.01623	99.00000	0.00000	-0.03269
V11	-0.03112	-0.00647	-0.00647	-0.02197	-0.01604	-0.01461	99.00000	-0.01978	-0.00647	99.00000	0.00000	-0.01303
V12	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V16	-0.04954	-0.01030	-0.01030	-0.03499	-0.02553	-0.02326	99.00000	-0.03150	-0.01030	99.00000	0.02655	-0.02075
V18	1.00000	-0.02195	-0.02195	-0.07453	-0.05439	-0.04954	99.00000	-0.06709	-0.02195	99.00000	0.00808	-0.04421
V20	-0.02195	1.00000	-0.00457	-0.01550	-0.01131	-0.01030	99.00000	-0.01396	-0.00457	99.00000	-0.03529	-0.00920
V21	-0.02195	-0.00457	1.00000	-0.01550	-0.01131	-0.01030	99.00000	-0.01396	-0.00457	99.00000	-0.10587	-0.00920
V23	-0.07453	-0.01550	-0.01550	1.00000	-0.03841	-0.03499	99.00000	-0.04738	-0.01550	99.00000	0.19968	-0.03122
V25	-0.05439	-0.01131	-0.01131	-0.03841	1.00000	-0.02553	99.00000	-0.03458	-0.01131	99.00000	0.03886	-0.02279
V26	-0.04954	-0.01030	-0.01030	-0.03499	-0.02553	1.00000	99.00000	-0.03150	-0.01030	99.00000	-0.01593	-0.02075
V27	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	1.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V29	-0.06709	-0.01396	-0.01396	-0.04738	-0.03458	-0.03150	99.00000	1.00000	-0.01396	99.00000	0.12383	-0.02810
V31	-0.02195	-0.00457	-0.00457	-0.01550	-0.01131	-0.01030	99.00000	-0.01396	1.00000	99.00000	-0.05882	-0.00920
V32	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	1.00000	99.00000	99.00000
V33	0.00808	-0.03529	-0.10587	0.19968	0.03886	-0.01593	99.00000	0.12383	-0.05882	99.00000	1.00000	0.10660
V39	-0.04421	-0.00920	-0.00920	-0.03122	-0.02279	-0.02075	99.00000	-0.02810	-0.00920	99.00000	0.10660	1.00000
V40	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V41	-0.02195	-0.00457	-0.00457	-0.01550	-0.01131	-0.01030	99.00000	-0.01396	-0.00457	99.00000	0.03529	-0.00920
V34	0.05312	-0.03570	0.12792	-0.02020	-0.08846	-0.00671	99.00000	-0.05354	0.12792	99.00000	-0.33336	-0.07189
V42	0.02199	0.03038	0.03038	0.10316	-0.07393	-0.09452	99.00000	0.09287	0.03038	99.00000	0.18405	0.06119
V43	0.08329	0.03949	0.03949	0.08618	-0.03029	0.01910	99.00000	0.01533	0.03949	99.00000	0.22887	0.07952
V44	0.12594	0.07823	-0.05837	-0.19815	0.08103	0.05325	99.00000	-0.17839	0.07823	99.00000	-0.41908	-0.11754
V45	0.05972	0.06398	0.06398	-0.15974	-0.06504	-0.03888	99.00000	-0.08023	0.06398	99.00000	-0.46913	-0.14372
V141	-0.06757	-0.04811	-0.04811	0.01324	0.05799	0.08509	99.00000	-0.00132	-0.04811	99.00000	0.27466	0.04715
V137	-0.03398	0.06639	0.15663	-0.11249	-0.06623	-0.01917	99.00000	-0.07292	-0.01217	99.00000	-0.41274	-0.10395
V158	-0.08787	-0.01828	-0.01828	-0.06206	-0.04529	-0.04125	99.00000	-0.05587	-0.01828	99.00000	0.20405	-0.03681
V159	-0.18983	-0.03949	-0.03949	-0.13406	-0.09785	-0.08911	99.00000	-0.12068	-0.03949	99.00000	-0.25431	-0.07952
V160	-0.11100	-0.02309	-0.02309	-0.07939	-0.05721	-0.05211	99.00000	-0.07057	-0.02309	99.00000	-0.00259	-0.04650
V161	-0.09401	-0.01955	-0.01955	-0.06639	-0.04846	-0.04413	99.00000	-0.05977	-0.01955	99.00000	0.02667	-0.03938
V162	-0.08469	-0.01762	-0.01762	-0.05981	-0.04365	-0.03976	99.00000	-0.05384	-0.01762	99.00000	-0.09724	-0.03546
V163	-0.08787	-0.01828	-0.01828	-0.06206	-0.04529	-0.04125	99.00000	-0.05587	-0.01828	99.00000	-0.00316	-0.03681

	V40	V41	V34	V42	V43	V44	V45	V141	V137	V158	V159	V160
V1	99.00000	0.06258	-0.32813	0.05228	0.04248	-0.19696	-0.23630	0.26069	-0.25993	0.09415	-0.06604	0.00470
V3	99.00000	0.13911	-0.17260	0.21842	0.17765	-0.41956	-0.30452	0.14376	-0.25711	0.41922	-0.20620	-0.05258
V4	99.00000	-0.08516	0.23836	-0.35679	-0.18508	0.68536	0.35187	-0.10688	0.19703	-0.22979	0.39938	0.14909
V5	99.00000	-0.03331	-0.12161	0.22163	0.05119	-0.42572	-0.13024	-0.01125	0.01289	-0.13333	-0.28802	-0.13095
V6	99.00000	-0.01550	-0.12119	0.04740	0.13406	-0.15599	-0.24729	0.05739	-0.08061	0.18617	-0.13406	-0.07839
V7	99.00000	-0.17342	0.14073	-0.10336	-0.22770	0.22791	0.41152	-0.06386	0.09895	-0.05452	0.22770	0.15314
V8	99.00000	0.22637	-0.07739	0.09989	0.17443	-0.15726	-0.31526	0.03319	-0.05803	-0.08075	-0.17443	-0.10200
V9	99.00000	-0.00457	-0.03570	-0.15028	-0.11564	0.07823	0.06398	-0.04811	0.17023	-0.01828	-0.03949	-0.02309
V10	99.00000	-0.01623	-0.12689	0.10800	0.14036	-0.08607	-0.13340	0.12555	0.00822	-0.06497	-0.14036	-0.08207
V11	99.00000	-0.00647	-0.05060	-0.08497	-0.05397	0.01408	-0.00523	0.03317	-0.10548	-0.02591	-0.05597	-0.03273
V12	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V16	99.00000	-0.01030	0.21483	-0.17606	-0.12094	0.05325	0.08331	-0.10857	-0.05121	-0.04125	-0.08911	-0.05211
V18	99.00000	-0.02195	0.05312	0.02199	0.08329	0.12594	0.05972	-0.06757	-0.03398	-0.08787	-0.18983	-0.11100
V20	99.00000	-0.00457	-0.03570	0.03038	0.03949	0.07823	0.06398	-0.04811	0.06639	-0.01828	-0.03949	-0.02309
V21	99.00000	-0.00457	0.12792	0.03038	0.03949	-0.05837	0.06398	-0.04811	0.15663	-0.01828	-0.03949	-0.02309
V23	99.00000	-0.01550	-0.02020	0.10316	0.08613	-0.19815	-0.15874	0.01324	-0.11249	-0.06206	-0.13406	-0.07839
V25	99.00000	-0.01131	-0.08544	-0.07393	-0.03029	0.08103	-0.06504	0.05799	-0.06623	-0.04529	-0.09785	-0.05721
V26	99.00000	-0.01030	-0.00671	-0.09452	0.01910	0.05325	-0.03888	0.08509	-0.01917	-0.04125	-0.08911	-0.05211
V27	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V29	99.00000	-0.01396	-0.05354	0.09287	0.01533	-0.17839	-0.08023	-0.00132	-0.07292	-0.05587	-0.12068	-0.07057
V31	99.00000	-0.00457	0.12792	0.03038	0.03949	0.07823	0.06398	-0.04811	-0.01217	-0.01828	-0.03949	-0.02309
V32	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V33	99.00000	0.03529	-0.33336	0.18405	0.22887	-0.41908	-0.46913	0.27466	-0.41274	0.20405	-0.25431	-0.00259
V39	99.00000	-0.00920	-0.07189	0.06119	0.07952	-0.11754	-0.14372	0.04713	-0.10395	-0.03681	-0.07952	-0.04658
V40	1.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000
V41	99.00000	1.00000	-0.03570	0.03038	0.03949	-0.05837	-0.07137	0.09492	-0.07442	-0.01828	-0.03949	-0.02309
V34	99.00000	.03570	1.00000	-0.11556	-0.12081	0.25565	0.32385	-0.30621	0.09715	-0.14290	0.09555	-0.00065
V42	99.00000	.03038	-0.11556	1.00000	0.32311	-0.52059	-0.10932	-0.06571	-0.10013	0.07342	-0.15572	-0.00524
V43	99.00000	0.03949	-0.12081	0.32311	1.00000	-0.14919	-0.53240	-0.04779	-0.11454	0.15807	-0.23345	-0.03906
V44	99.00000	-0.05837	0.25565	-0.52059	-0.14919	1.00000	0.41293	-0.14816	0.21537	-0.19719	0.29685	0.03522
V45	99.00000	-0.07137	0.32385	-0.10932	-0.53240	0.41293	1.00000	-0.17378	0.20304	-0.17732	0.34629	-0.09306
V141	99.00000	0.09492	-0.30621	-0.06571	-0.04779	-0.14816	-0.17378	1.00000	-0.12921	0.11277	-0.12890	0.10262
V137	99.00000	-0.07442	0.09715	-0.10013	-0.11454	0.21537	0.20304	-0.12921	1.00000	-0.11479	0.21458	-0.02518
V158	99.00000	-0.01828	-0.14290	0.07342	0.15807	-0.19719	-0.17732	0.11277	-0.11479	1.00000	-0.15807	-0.09243
V159	99.00000	-0.03949	0.09555	-0.15572	-0.23345	0.29685	0.34629	-0.12890	0.21458	-0.15807	1.00000	-0.19967
V160	99.00000	-0.02309	-0.00065	-0.00524	-0.03906	0.03522	-0.09306	0.10262	-0.02518	-0.09243	-0.19967	1.00000
V161	99.00000	-0.01955	0.01199	0.08461	0.01279	-0.21554	0.00124	-0.06191	-0.07561	-0.07828	-0.16910	-0.09888
V162	99.00000	-0.01762	0.08771	-0.08192	-0.01865	0.18890	0.13495	-0.06736	0.13170	-0.07052	-0.15234	-0.08908
V163	99.00000	-0.01828	-0.05557	0.12163	0.03387	-0.16073	-0.14120	0.03643	0.01123	-0.07317	-0.15807	-0.09243

	V161	V162	V163
V1	-0.01892	-0.21397	0.06174
V3	-0.14057	-0.12663	-0.03963
V4	-0.29478	0.20686	-0.26683
V5	0.50127	-0.12849	0.36691
V6	-0.06639	-0.05981	0.18617
V7	0.01212	0.10158	-0.26775
V8	0.03783	-0.07782	0.18243
V9	-0.01955	-0.01762	-0.01828
V10	-0.06951	-0.06262	-0.06497
V11	-0.02772	-0.02497	-0.02591
V12	99.00000	99.00000	99.00000
V16	-0.04413	-0.03976	-0.04125
V18	-0.09401	-0.08469	-0.08787
V20	-0.01955	-0.01762	-0.01828
V21	-0.01955	-0.01762	-0.01828
V23	-0.06639	-0.05981	-0.06206
V25	-0.04846	-0.04365	-0.04529
V26	-0.04413	-0.03976	-0.04125
V27	99.00000	99.00000	99.00000
V29	-0.05977	-0.05384	-0.05587
V31	-0.01955	-0.01762	-0.01828
V32	99.00000	99.00000	99.00000
V33	0.02667	-0.09724	-0.00314
V39	-0.03938	-0.03548	-0.03681
V40	99.00000	99.00000	99.00000
V41	-0.01955	-0.01762	-0.01828
V34	0.01199	0.08771	-0.05557
V42	0.08461	-0.08192	0.12163
V43	0.01279	-0.01865	0.03387
V44	-0.21554	0.18890	-0.16073
V45	0.00124	0.13495	-0.14120
V141	-0.06191	-0.06736	0.03643
V137	-0.07561	0.13170	0.01123
V158	-0.07828	-0.07052	-0.07317
V159	-0.16910	-0.15234	-0.15807
V160	-0.09888	-0.08908	-0.09243
V161	1.00000	-0.07544	-0.07828
V162	-0.07544	1.00000	-0.07052
V163	-0.07828	-0.07052	1.00000